



Høgskolen i Molde
Vitenskapelig høyskole i logistikk

1. TESTeksamen

Løsning

“MAT110 Statistikk I”, 2021
versjon 02

- mai 2021

Oppgave 1 — 30%

Logistikk- og budfirmaet JetPak leverer blant annet pakker fra dør til dør.

De ansatte i firmaet ønsker å se nærmere på pakker som blir skadet eller forsinket. Derfor definerer de følgende begivenheter:

$$S = \text{pakke som blir skadet under transport} \quad (1.1)$$

$$F = \text{pakke som blir forsinket ved transport} \quad (1.2)$$

Det er da totalt 4 forskjellige kombinasjoner av begivenheter som pakkene kan deles inn i:

$$\bar{S} \cap \bar{F}, \bar{S} \cap F, S \cap \bar{F}, S \cap F \quad (1.3)$$

Anta at JetPak leverer $N = 100$ pakker i løpet av en dag.

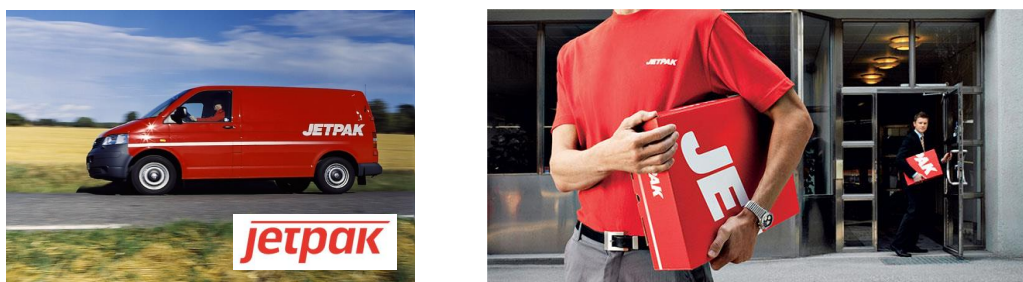
Resultatet av disse leveransene er vist i figur 1.1.

I denne oppgaven skal vi bruke dette som grunnlag for beregningene.

begivenhet	$\bar{S} \cap \bar{F}$	$\bar{S} \cap F$	$S \cap \bar{F}$	$S \cap F$
antall pakker	70	15	10	5

Figur 1.1: Pakker.

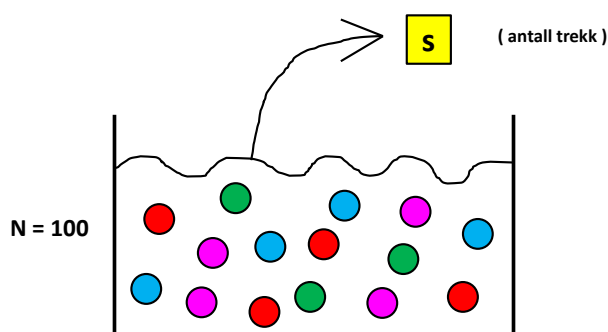
Anta at alle begivenhetene S og F er uavhengige av hverandre.



Figur 1.2: JetPak.

La oss beskrive problemet til JetPak med en **urne**.

Anta at vi trekker tilfeldige pakker fra urnen og at det er lik sannsynlighet for å trekke de forskjellige pakkene.



Figur 1.3: Urne.

- a) (2.5%) Hvilke to betingelser må være oppfylt for at urnemodellen skal gjelde? Gjelder urnemodellen for vårt tilfelle med pakker?

Løsning:

Betingelsene

- **uniformt** utfallsrom
- utfallene ikke påvirker hverandre, dvs. de er **uavhengige**

må være oppfylte for at urnmodellen skal gjelde.

I oppgaven antar man at pakkene er uavhengige av hverandre.

I tillegg antas det i oppgaven at det er lik sannsynlighet for å trekke de forskjellige pakkene fra urnen, altså uniformt utfallsrom.

Ja, urnemodellen gjelder for vårt tilfelle med pakker.

Ett trekk ($s = 1$)

La oss først se på problemet når vi kun trekker – altså leverer – en tilfeldig valgt pakke av de $N = 100$ pakkene.

- b)** (2.5%) Vis – ved svært korte utregninger basert på tallene i figur 1.1 – at vi har følgende sannsynligheter:

$$P(\bar{S} \cap \bar{F}) = 0.70 \quad (1.4)$$

$$P(\bar{S} \cap F) = 0.15 \quad (1.5)$$

$$P(S \cap F) = 0.05 \quad (1.6)$$

Løsning:

Siden urnemodellen gjelder så finner vi sannsynlighetene via telling, dvs. #gunstige/#mulige. Med kun ett trekk ($s = 1$) så finner man sannsynlighetene ved svært korte utregninger:

$$P(\bar{S} \cap \bar{F}) = \frac{\# g}{\# m} = \frac{70}{100} = 0.70 \quad (1.7)$$

$$P(\bar{S} \cap F) = \frac{\# g}{\# m} = \frac{15}{100} = 0.15 \quad (1.8)$$

$$P(S \cap \bar{F}) = \frac{\# g}{\# m} = \frac{10}{100} = 0.10 \quad (1.9)$$

$$P(S \cap F) = \frac{\# g}{\# m} = \frac{5}{100} = 0.05 \quad (1.10)$$

JetPak taper penger på pakker som blir skadet eller forsinket.

For å se nærmere på utgiftene forbundet med disse uønskede begivenhetene så introduseres en stokastisk variabel U :

$$U = \text{utgift for JetPak forbundet med skade eller forsinkelse av en pakke} \quad (1.11)$$

Dersom en pakke kommer frem uskadet til rett tid så har ikke JetPak noen utgift.

Anta videre at utgiften til JetPak er 500 NOK dersom en pakke er forsinket, men ikke skadet.

Dersom en pakke er skadet, men ikke forsinket, så er utgiften 900 NOK.

Og – til slutt – dersom en pakke er både skadet og forsinket så er utgiften 1 500 NOK.

Utfallsrommet til U er dermed:

$$\Omega = \{ 0, 500, 900, 1500 \} \quad (1.12)$$

- c) (2.5%) Hva er forventet utgift $E[U]$ for JetPak?

Løsning:

Sammenhengen mellom lign.1.7-1.10 og $P(U)$, er:

$$P(U = 0) = P(\bar{S} \cap \bar{F}) \quad (1.13)$$

$$P(U = 500) = P(\bar{S} \cap F) \quad (1.14)$$

$$P(U = 900) = P(S \cap \bar{F}) \quad (1.15)$$

$$P(U = 1500) = P(S \cap F) \quad (1.16)$$

Bruker disse resultatene og regner ut JetPak sine forventede utgifter per pakke:

$$\begin{aligned} E[U] &= \sum_{i=1}^4 u_i \cdot P(U = u_i) \\ &= \left(0 \cdot \overbrace{P(\bar{S} \cap \bar{F})}^{= 0.70} + 500 \cdot \overbrace{P(\bar{S} \cap F)}^{= 0.15} + 900 \cdot \overbrace{P(S \cap \bar{F})}^{= 0.10} + 1500 \cdot \overbrace{P(S \cap F)}^{= 0.05} \right) \text{ NOK} \\ &= 240 \text{ NOK} \end{aligned} \quad (1.17)$$

- d) (2.5%) Figur 1.4 viser Venn-diagrammet for de 4 aktuelle kombinasjonene av begivenheter.

Er disse 4 kombinasjonene av begivenheter **disjunkte**?
Gi en kort begrunnelse.

Løsning:

Fra figuren i oppgaver ser vi at de 4 aktuelle kombinasjonene av begivenheter ikke overlapper. Derfor er de disjunkte. ¹

$P((\bar{S} \cap F) \cup (S \cap \bar{F}))$ er sannsynligheten for at en tilfeldig valgt pakke er enten bare skadet eller bare forsinket.

- e) (2.5%) Vis at sannsynligheten for at en tilfeldig valgt pakke er enten bare skadet eller bare forsinket, er:

$$P((\bar{S} \cap F) \cup (S \cap \bar{F})) = 0.25 \quad (1.18)$$

Løsning:

Bruker addisjonssetningen og resultater fra oppgave 1a):

$$\begin{aligned} P((\bar{S} \cap F) \cup (S \cap \bar{F})) &= \overbrace{P(\bar{S} \cap F)}^{= 0.10} + \overbrace{P(S \cap \bar{F})}^{= 0.15} - \overbrace{P((\bar{S} \cap F) \cap (S \cap \bar{F}))}^{= 0 \text{ (disjunkt)}} \\ &= 0.10 + 0.15 = 0.25, \quad \text{q.e.d.} \end{aligned} \quad (1.19) \quad (1.20)$$

Merk: Fra oppgave 1d) vet vi at $\bar{S} \cap F$ og $S \cap \bar{F}$ er disjunkte.

¹Ekstrakommentar som man ikke behøver å ha med på eksamen:
I dette tilfellet overlapper de akkurat ikke - i analogi med et puslespill.

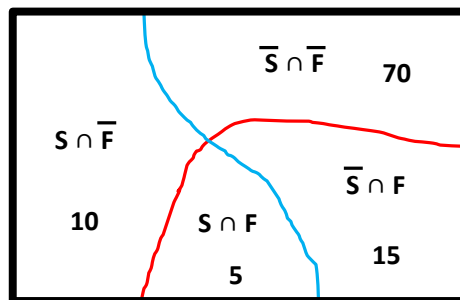
- f) (2.5%) Finn samme sannsynlighet som i oppgave 1e) ved å bruke enkel kombinatorikk og tallene i figur 1.1 direkte.²

Løsning:

Med kun ett trekk ($s = 1$) så finner man den aktuelle sannsynligheten ved en svært kort utregning:

$$P((\bar{S} \cap F) \cup (S \cap \bar{F})) = \frac{\# g}{\# m} = \frac{15 + 10}{100} = 0.25 \quad (1.22)$$

altså sammen sannsynlighet som i oppgave 1e) – slik som det skal være.



Figur 1.4: Venn-diagram.

²I likhet med oppgave 1b) så er utregningen i denne oppgaven svært kort. Bruk grunnleggende kombinatorikk:

$$P((\bar{S} \cap F) \cup (S \cap \bar{F})) = \frac{\# \text{ gunstige kombinasjoner}}{\# \text{ mulige kombinasjoner}} \quad (1.21)$$

og tallene i figur 1.1.

- g) (2.5%) Hva er sannsynligheten for at en tilfeldig valgt pakke er skadet?

Løsning:

Oppsplitting av utfallsrom:

$$P(S) = P(S \cap F) + P(S \cap \bar{F}) = 0.05 + 0.10 = 0.15 \quad (1.23)$$

- h) (2.5%) Man kan vise at sannsynligheten for at en tilfeldig valgt pakke er uskadet eller kommer til riktig tid, er:

$$P(\bar{S} \cup \bar{F}) = 0.95 \quad (1.24)$$

(Dette faktum skal du ikke vise. Bare ta det for gitt.) ³

Bruk lign.(1.24) og den ene tvillingsetningen ⁴ til å finne sannsynligheten for at en tilfeldig valgt pakke er både skadet og forsinket, dvs. finn $P(S \cap F)$.

Stemmer denne sannsynligheten med oppgave 1b)?

Gi en *kort* begrunnelse.

Løsning:

Bruker den ene tvillingsetningen samt den oppgitte sannsynligheten:

$$\begin{aligned} P(S \cap F) &= 1 - P(\bar{S} \cup \bar{F}) \\ &= 1 - 0.95 = 0.05 \end{aligned} \quad (1.25)$$

som stemmer med resultatet fra oppgave 1b).

Dette er slik det må være siden vi i oppgavene 1b) og 1h) regner ut samme sannsynlighet med to forskjellige metoder.

³Selv om du ikke blir bedt om å vise lign.(1.24) så kan du gjerne prøve likevel siden dette er en TESTeksamen. På den måten får du mer innblikk i situasjonen.

⁴Se gjerne boken.

Flere trekk ($s > 1$)

La oss nå se på problemet når vi trekker — altså leverer — flere (dvs. flere enn én) tilfeldig valgte pakker av de $N = 100$ pakkene. Rekkefølgen pakkene trekkes i betyr ikke noe.

- i) (2.5%) I dette kurset har vi lært om 4 forskjellige situasjoner i forbindelse med kombinatorikk. Gi en kort begrunnelse for hvorfor dette — å trekke fra de utvalgte pakkene — tilsvarer situasjon 3.⁵

Løsning:

I oppgaven er det oppgitt at:

- Rekkefølgen pakkene trekkes i betyr ikke noe
- Når en pakke er trykket så legger man den ikke tilbake i urnen, dvs. uten tilbakelegging

Dermed tilsvarer dette situasjon 3.

- j) (7.5%) Anta at JetPak leverer $s = 12$ tilfeldige pakker en dag.

Hva er sannsynligheten P_{mix} for at JetPak leverer 3 pakker fra hver av de 4 aktuelle begivenhetene i figur 1.1?⁶

Løsning:

Sannsynligheten P_{mix} for at Jetpak leverer 3 pakker fra hver av de 4 aktuelle begivenhetene:

⁵Se gjerne boken angående hva situasjon 3 er.

⁶Denne oppgaven er litt vanskelig. Dersom dere greier litt så får dere poeng. Prøv å ikke lever inn blankt.

$$P_{\text{mix}} = \frac{\# g}{\# m} \quad (1.26)$$

$$= \frac{(\# \text{ komb. } S \cap \bar{F}) \cdot (\# \text{ komb. } \bar{S} \cap F) \cdot (\# \text{ komb. } S \cap F) \cdot (\# \text{ komb. } \bar{S} \cap \bar{F})}{\text{totalt}} \quad (1.27)$$

$$= \frac{\binom{10}{3} \cdot \binom{15}{3} \cdot \binom{5}{3} \cdot \binom{70}{3}}{\binom{100}{12}} = 2.85 \cdot 10^{-5} \quad (1.28)$$

altså svært liten sannsynlighet.⁷

■

⁷Legg merke til at “summeregelen” er opfylt i lign.(1.28).

Oppgave 2 — 20%

Transportselskapet *Fjord1* har sett nærmere på antall personbiler som ankommer Molde fergekai. *Fjord1* har funnet ut at det i gjennomsnitt kommer 920 personbiler alle hverdager i tidsrommet 10:00 - 14:00, altså i et tidsrom på 4 timer. Anta, for enkelhets skyld, at det kun er personbiler som ankommer fergekaien i dette tidsrommet, og at raten per tid er konstant. Anta også at alle bilene som ankommer fergekaien, er uavhengige av hverandre.

Det kommer en ny ferge **hver halvtime** i det aktuelle tidsrommet. Derfor definerer vi følgende stokastiske variabel:

$$X = \text{antall personbiler som ankommer fergekaien i en 30 minutters periode,} \\ \text{dvs. mellom to ferger} \quad (1.29)$$

- a) (2.5%) Hva slags diskret sannsynlighetsfordeling har den stokastiske variabelen X ?

Løsning:

Siden denne oppgaven dreier seg om antall begivenheter innenfor et gitt tidsintervall (30 minutter), altså en rate, så vil den stokastiske variabelen X beskrives av en ⁸

$$\text{Poissonfordeling} \quad (1.30)$$

- b) (2.5%) Vis, ved en kort utregning, at raten λ som beskriver antall biler per halvtime er

$$\lambda = 115 \quad (1.31)$$

⁸Her er ikke “suksess”-sannsynligheten p og antall forsøk n oppgitt. Så en binomisk sannsynlighetsfordeling er ikke hensiktsmessig for situasjonen som beskrevet i oppgaven.

Løsning:

I oppgaveteksten opplyses det at det i gjennomsnitt kommer 920 personbiler per 4 timer til fergekaia. Antall biler som kommer hver halvtime λ er derfor:

$$\lambda = 920 \cdot \frac{0.5 \text{ time}}{4 \text{ timer}} = 115, \text{ q.e.d.} \quad (1.32)$$

- c) (2.5%) i) Hva er forventet antall biler $E[X]$ som ankommer fergekaia mellom to ferger?

ii) Hva er standardavviket $\sigma[X]$ av antall biler som ankommer fergekaia mellom to ferger?

Løsning:

i) Fra oppgave 2a) og 2b) vet vi at $X \sim \text{Poi}[\lambda]$, hvor $\lambda = 115$.
Forventet antall biler $E[X]$ som ankommer fergekaia mellom to ferger blir derfor:⁹

$$E[X] = \lambda = 115 \quad (1.33)$$

ii) Standardavviket $\sigma[X]$ av antall biler som ankommer fergekaia mellom to ferger er da:

$$\sigma[X] = \sqrt{\text{Var}[X]} = \sqrt{\lambda} = \sqrt{115} = 10.72 \quad (1.34)$$

⁹Se læreboken.



Figur 1.5: Ferge.

- d) (2.5%) i) En diskret Poisson-sannsynlighetsfordeling $\text{Poi}[\lambda]$ kan, under bestemte betingelser, med god tilnærming beskrives av en kontinuerlig sannsynlighetsfordeling.

Hvilken sannsynlighetsfordeling?

- ii) Hva slags betingelse må raten λ oppfylle for at tilnærmelsen som beskrevet ovenfor skal gjelde?¹⁰

- iii) Er denne betingelsen på λ fra oppgave ii oppfylt for vårt tilfelle?

Løsning:

- i) En diskret Poisson sannsynlighetsfordelingen $\text{Poi}[\lambda]$ kan, under bestemte betingelser, med god tilnærming beskrives av en normalfordeling:

$$\text{Poi}[\lambda] \longrightarrow N[\mu = \lambda, \sigma = \sqrt{\lambda}] \quad (1.35)$$

- ii) I læreboken ser vi at dersom

$$\lambda \gtrsim 5, \quad (1.36)$$

så gjelder påstanden fra oppgave 2d) i.

- iii) Siden $\lambda = 115 \gg 5$ så er betingelsen i lign.(1.36) oppfylt og tilnærmelsen i lign.(1.35) gjelder i vårt tilfelle.

¹⁰Se læreboken.

Fergene som går strekningen Molde - Vestnes har alle en kapasitet på $X_0 = 125$ personbiler.

- e) (5%) Anta at det i utgangspunktet er 20 biler i fergekø i tillegg til de bilene som ankommer fergekaia med en konstant rate λ .

Hva er da sannsynligheten for at ikke alle bilene får plass i fergen?

Løsning:

Ved å bruke resultatet fra oppgave 2d) i, dvs.

$$X \sim \text{Poi}[\lambda] \longrightarrow N[\mu = \lambda, \sigma = \sqrt{\lambda}] \quad (1.37)$$

så kan vi finne sannsynligheten ved standardisering og deretter tabelloppslag.

Siden 20 biler i fergekø så er det altså ikke alle bilene som får plass i fergen.

(Oppgitt i oppgaven: Fergene har kapasitet på $X_0 = 125$ biler.)

$$P(X > X_0 - 20) = P(X > 125 - 20) \quad (1.38)$$

$$= 1 - P(X \leq 105) \quad (1.39)$$

$$\stackrel{\text{standardiser}}{=} 1 - P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{=Z} \leq \frac{105 - E[X]}{\sigma[X]}\right) \quad (1.40)$$

$$= 1 - P\left(Z \leq \frac{105 - \lambda}{\sqrt{\lambda}}\right) \quad (1.41)$$

$$= 1 - P\left(Z \leq \frac{105 - 115}{\sqrt{115}}\right) \quad (1.42)$$

$$= 1 - P(Z \leq -0.93) \quad (1.43)$$

$$= 1 - \left(1 - P(Z \leq 0.93)\right) = P(Z \leq 0.93) \quad (1.44)$$

Dermed:

$$P(X > X_0 - 20) = P(Z \leq 0.93) \quad (1.45)$$

$$= \underbrace{G(0.93)}_{\text{se tabell}} = 0.8238 \quad (1.46)$$

Dersom det i utgangspunktet er 20 biler i fergekø i tillegg de til som ankommer fergekaia med konstant rate λ så er det 82.38 % sannsynlighet for at ikke alle bilene får plass i fergen.

- f) (5%) Anta nå at det i utgangspunktet *ikke* er noen biler i fergekø.

Dersom *Fjord 1* ønsker at det skal være 95 % sannsynlighet for at alle bilene kommer med, hvor stor kapasitet X_{kap} må fergene ha da?

Løsning:

Fjord 1 ønsker at det skal være 95 % sannsynlighet for at alle bilene kommer med. La X_{kap} være den ukjente kapasiteten til fergen som vi ønsker å finne. Denne er bestemt av:

$$P(X \leq X_{\text{kap}}) = 0.95 \quad (1.47)$$

$$P\left(\underbrace{\frac{X - E[X]}{\sigma[X]}}_{\equiv Z} \leq \underbrace{\frac{X_{\text{kap}} - E[X]}{\sigma[X]}}_{\equiv Z_{\text{kap}}}\right) \stackrel{\text{standardiser}}{=} 0.95 \quad (1.48)$$

$$P(Z \leq Z_{\text{kap}}) = 0.95 \quad (1.49)$$

Ved “**omvendt tabelloppslag**” ser vi at svaret 0.95 ligger midt mellom argumentene $Z_{\text{kap}} = 1.64$ og $Z_{\text{kap}} = 1.65$. Dermed:

$$Z_{\text{kap}} = 1.645 \quad (1.50)$$

Via

$$Z_{\text{kap}} = \frac{X_{\text{kap}} - E[X]}{\sigma[X]} \quad (1.51)$$

så løser vi med hensyn på X_{kap} alene:

$$X_{\text{kap}} = Z_{\text{kap}} \cdot \overbrace{\sigma[X]}^{= \sqrt{\lambda}} + \overbrace{E[X]}^{= \lambda} \quad (1.52)$$

$$= Z_{\text{kap}} \cdot \sqrt{\lambda} + \lambda \quad (1.53)$$

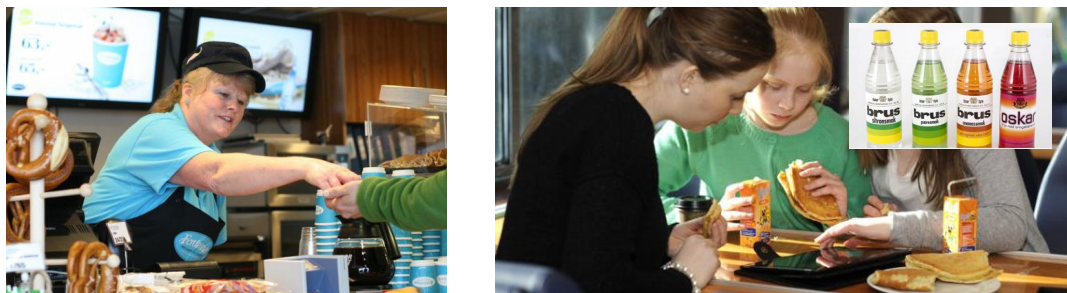
$$= 1.645 \cdot \sqrt{115} + 115 = 132.64 \approx 133 \quad (1.54)$$

For at det skal være 95 % sikkert at alle i fergekøen skal komme med fergen så må fergen ha en kapasitet på 133 personbiler .

■

Oppgave 3 — 30%

I tillegg til fergebilletter er også kiosksalg en viktig inntektskilde for *Fjord1*. I den sammenheng skal vi se nærmere på salg av brus. Vi skal bruke to forskjellige statistiske tilnærmelser, modell 1 og modell 2.



Figur 1.6: Kiosk.

Modell 1:

For modell 1 defineres følgende stokastiske variabel:

Y = antall passasjerer som kjøper **en** brus på en gitt overfart

Basert på historiske data har *Fjord1* funnet ut at det er **11 %** av fergepassasjerene som kjøper en brus i løpet av en gitt overfart. Anta at fergepassasjerene kjøper brus uavhengige av hverandre. Anta også at det i gjennomsnitt er 3 passasjerer i hver bil, at det i utgangspunktet ikke er noen biler i fergekø og at alle bilene som ønsker å bli med fergen får plass. Selv om antall ankomne biler egentlig er en stokastisk størrelse så antas det i denne oppgaven at det faktisk kommer $\lambda = 115$ antall biler til fergekaia mellom to fergeravganger. (Parameteren λ er den samme som i oppgave 2).

- a) (2.5%) Begrunn hvorfor det er rimelig å anta at Y er binomisk fordelt, dvs. begrunn hvorfor¹¹

$$Y \sim \text{Bin}[n = 3\lambda, p = 0.11] \quad (1.55)$$

Kommenter *kort* hvorfor $n = 3\lambda$.

¹¹Hvilke 4 forutsetninger må være oppfylt for at Y skal være binomisk fordelt? Er disse forutsetningene oppfylt i vårt tilfelle? (Se læreboken).

Løsning:

4 forutsetninger må være oppfylt for at Y skal være binomisk fordelt:

1. Hvert forsøk skal ha 2 mulige utfall, s (suksess) eller f (fiasko):
Enten så kjøper en passasjer en brus, eller så gjør han/hun det ikke.
2. Det skal være samme sannsynlighet $p = 0.11$ for suksess i alle n forsøkene.
3. Alle forsøk er uavhengige:
Fergepassasjerene kjøper brus uavhengige av hverandre.
4. Vi gjennomfører et bestemt antall forsøk, n :
Fra oppgave 2d) vet vi at det kommer $\lambda = 115$ biler hver halvtime, dvs. mellom to ferger.
Selv om antall ankomne biler egentlig er en stokastisk størrelse så antas det i denne oppgaven at det “faktisk kommer λ antall biler til fergekaien mellom to fergeravganger”.
I oppgaven får vi også opplyst at det i utgangspunktet ikke er noen biler i fergekø.
I tillegg får vi opplyst at det i gjennomsnitt er 3 passasjerer i hver bil.
Derfor er: $n = 3\lambda = 3 \cdot 115 = 345$ er et bestemt antall “forsøk”.

Alle de 4 forutsetningene for en binomisk fordeling er oppfylt. Derfor er det rimelig å anta at Y er binomisk fordelt, dvs. $Y \sim \text{Bin}[n = 3\lambda = 345, p = 0.11]$.

- b) (5%) Finn forventet antall solgte brus på en gitt overfart ifølge modell 1, dvs. finn $E[Y]$.

Løsning:

Siden $Y \sim \text{Bin}[n = 3\lambda = 345, p = 0.11]$ så er forventet antall solgte brus på en gitt overfart ifølge modell 1:

$$E[Y] \stackrel{\text{Bin.}}{=} n \cdot p = 345 \cdot 0.11 = 37.95 \quad (1.56)$$

- c) (2.5%) i) Finn variansen til antall solgte brus for en overfart ifølge modell 1, dvs. finn $\text{Var}[Y]$.
- ii) Hva er det tilhørende standardavviket $\sigma[Y]$?

Løsning:

i) Variansen til antall solgte brus på en gitt overfart ifølge modell 1:

$$\text{Var}[Y] \stackrel{\text{Bin.}}{=} n \cdot p \cdot (1 - p) = 345 \cdot 0.11 \cdot (1 - 0.11) = 33.78 \quad (1.57)$$

ii) Tilhørende standardavviket $\sigma[X]$:

$$\sigma[X] = \sqrt{\text{Var}[X]} = \sqrt{33.78} = 5.81 \quad (1.58)$$

- d) (2.5%) i) Dersom en betingelse er oppfylt så kan en binomisk fordeling tilnærmes med en normalfordeling. Hvilken betingelse er det? ¹²
- ii) Er denne betingelsen oppfylt i vårt tilfelle?

¹²Se læreboken.

Løsning:

i) Betingelse som må være oppfylt så for at en binomisk fordeling kan tilnærmes med en normalfordeling er: ¹³

$$n \cdot p(1-p) \gtrsim 5 \quad (1.59)$$

ii) For vårt tilfelle:

$$345 \cdot 0.11(1-0.11) = 33.78 \gg 5 \quad (1.60)$$

Ja, betingelsen er godt oppfylt for vårt tilfelle.

- e) (5%) Hva er sannsynligheten for at det selges mer enn 45 brus på en overfart ifølge modell 1?

Løsning:

For å regne ut sannsynligheten for at det selges mer enn 45 brus på en overfart så benytter vi at $Y \sim \text{Bin}[n, p]$ kan tilnærmes med en normalfordeling. Dermed kan vi standardisere og til slutt gjøre et tabellopslag.

$$P(Y > 45) = 1 - P(Y \leq 45) \quad (1.61)$$

$$= 1 - P\left(\underbrace{\frac{Y - E[Y]}{\sigma[Y]}}_{=Z} \leq \frac{45 - E[Y]}{\sigma[Y]}\right) \quad (1.62)$$

$$= 1 - P\left(Z \leq \frac{45 - 38}{5.81}\right) \quad (1.63)$$

$$= 1 - P(Z \leq 1.21) \quad (1.64)$$

$$= 1 - \underbrace{G(1.21)}_{\substack{\text{tabell} \\ = 0.8869}} \quad (1.65)$$

$$= 1 - 0.8869 = 0.1131 \quad (1.66)$$

¹³Se gjerne læreboken.

Modell 2:

I modell 1 beskrevet ovenfor kjøper kunden enten en brus eller ingen brus. Dette er en forenkling av virkeligheten siden en kunde kan kjøpe mer enn en brus på en gitt overfart. Modellen som vi nå skal introdusere, modell 2, tar høyde for at fergepassasjerer kan kjøpe mer enn en brus. For denne modellen defineres derfor følgende stokastiske variabel:

B = antall brus som kjøpes *per passasjer* på en gitt overfart

Sannsynlighetsfordelingen $P(B = b_i)$, hvor $i = 0, 1, 2, 3$, er gitt ved:

b_i	0	1	2	3
$P(B=b_i)$	0.93	0.04	0.02	0.01

Figur 1.7: Sannsynlighetsfordeling $P(B = b_i)$.

Tilhørende forventning og standardavvik er:

$$E[B] = 0.11 \quad , \quad \sigma[B] = 0.4449 \quad (1.67)$$

(Du trenger ikke regne ut disse størrelsene. Bare ta dem for gitt).

- f) (5%) Hva er sannsynligheten at en tilfeldig valgt passasjer kjøper mer enn en brus?

Løsning:

Sannsynligheten at en tilfeldig valgt passasjer kjøper mer enn en brus:

$$P(B > 1) = P(B = 2) + P(B = 3) \quad (1.68)$$

$$= 0.02 + 0.01 = 0.03 \quad (1.69)$$

Alternativt kan denne løses på følgende måte:

$$P(B > 1) = 1 - P(B \leq 1) \quad (1.70)$$

$$= 1 - \left(P(B = 0) + P(B = 1) \right) \quad (1.71)$$

$$= 1 - 0.93 - 0.04 = 0.03 \quad (1.72)$$

Det er nok at oppgaven løses på en måte.

Det *totale* antallet brus som selges på en gitt overfart er:

$$B_{\text{tot}} = \sum_{i=1}^n B_i = B_1 + B_2 + \dots + B_n \quad (1.73)$$

hvor $n = 3\lambda$ er antall passasjerer på en gitt overfart.

g) (2.5%) Finn $E[B_{\text{tot}}]$.

Løsning:

Forventet antall brus solgt totalt i løpet av en overfart: ($n = 3\lambda = 345$)

$$E[B_{\text{tot}}] = E[B_1 + B_2 + \dots + B_n] \quad (1.74)$$

$$\stackrel{\text{alltid}}{=} \underbrace{E[B_1] + E[B_2] + \dots + E[B_n]}_{n = 345} = n \cdot \underbrace{E[B]}_{=0.11} = 345 \cdot 0.11 = 37.95 \quad (1.75)$$

NB: Overgangen i lign.(1.74) til (1.75) gjelder alltid. Uansett om de stokastiske variablene B_i er uavhengige eller ikke.

Anta at passasjerene kjøper brus uavhengige av hverandre.

h) (5%) Finn $\text{Var}[B_{\text{tot}}]$.

Løsning:

Variansen til antall brus solgt totalt i løpet av en overfart: ($n = 3\lambda = 345$)

$$\text{Var}[B_{\text{tot}}] = \text{Var}[B_1 + B_2 + \dots + B_n] \quad (1.76)$$

$$\stackrel{\text{uavhengig}}{=} \text{Var}[B_1] + \text{Var}[B_2] + \dots + \text{Var}[B_n] \quad (1.77)$$

$$= n \cdot \underbrace{\text{Var}[B]}_{= \sigma^2[B]} = n \cdot \sigma^2[B] = 345 \cdot 0.4449^2 = 68.28 \quad (1.78)$$

NB: Overgangen i lign.(1.76) til (1.77) gjelder kun dersom de stokastiske variablene B_i er uavhengige og dermed er kovariansen lik null. ¹⁴

- i) (2.5%) Hvilken fordeling er det rimelig å anta at B_{tot} har? Begrunn svaret.

Løsning:

Siden

1. fergepassasjerene kjøper brus uavhengig av hverandre:
 $B_i \sim$ er uavhengige for alle $i = 1, 2, 3, \dots, n$
2. alle bruskjøpere har samme sannsynlighetsfordeling for B_i :
 $B_i \sim$ samme sannsynlighetsfordeling for alle $i = 1, 2, 3, \dots, n$
3. antall “forsøk”, dvs. potensielle bruskjøpere, $n = 3\lambda = 345$, er tilstrekkelig stort ¹⁵

så gjelder **sentralgrensesetningen**. Dermed er B_{tot} normalfordelt med god tilnærming:

$$B_{\text{tot}} \sim N[E[B_{\text{tot}}], \text{Var}[B_{\text{tot}}]] \quad (1.79)$$

- j) (5%) Hva er sannsynligheten, ifølge modell 2, for at det selges mer enn 45 brus på en overfart?

Løsning:

For å regne ut sannsynligheten for at det selges mer enn 45 brus på en overfart så benytter vi at B_{tot} kan tilnærmes med en normalfordeling. Dermed løser man oppgaven med standardisering og til slutt tabelloppslag.

¹⁴Å poengtere dette må til for få 100% uttelling på oppgaven.

¹⁵Husk: Antall forsøk n for at sentralgrensesetningen skal gjelde er avhengig av situasjonen. Men en **tommelfingerregel** er at vi bør ha $n \gtrsim 30$.

$$P(B_{\text{tot}} > 45) = 1 - P(B_{\text{tot}} \leq 45) \quad (1.80)$$

$$= 1 - P\left(\underbrace{\frac{B_{\text{tot}} - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}}_{= Z_{\text{tot}}} \leq \frac{45 - E[B_{\text{tot}}]}{\sigma[B_{\text{tot}}]}\right) \quad (1.81)$$

$$= 1 - P(Z_{\text{tot}} \leq \frac{45 - 37.95}{\sqrt{68.28}}) \quad (1.82)$$

$$= 1 - P(Z_{\text{tot}} \leq 0.85) \quad (1.83)$$

$$= 1 - \underbrace{G(0.85)}_{\text{se tabell}} \quad (1.84)$$

$$= 1 - 0.8023 = 0.1977 \quad (1.85)$$

- k) (2.5%) Sammenlign og kommenter svarene fra oppgave 3c) i og 3h), dvs. sammenlign og kommenter $\text{Var}[Y]$ og $\text{Var}[B_{\text{tot}}]$.¹⁶

Løsning:

I modell 1 kjøper fergepassasjerene enten ingen brus eller en brus.

I modell 2, derimot, tas det høyde for at fergepassasjerene kan kjøpe mer enn en brus.

Derfor er det rimelig at variansen til modell 2 er større enn variansen til modell 1:

$$\underbrace{\text{Var}[B_{\text{tot}}] = 68.28}_{\text{modell 2}} > \underbrace{\text{Var}[Y] = 33.78}_{\text{modell 1}} \quad (1.86)$$

Anta at prisen på brus er $p = 20$ NOK. Innkjøpskostnaden er $k = 6$ NOK. Anta videre at

¹⁶Er det rimelig at den ene variansen er større enn den andre?

Fjord1 sine faste kostnader forbundet med kioskdirften som er relatert til brussalget er $k_{\text{fast}} = 175$ NOK. Fortjenesten F på brussalget for en gitt overfart er da gitt ved:

$$F = (p - k)B_{\text{tot}} - k_{\text{fast}} \quad (1.87)$$

- 1) (2.5%) Finn forventet fortjeneste $E[F]$ på brussalget til *Fjord1* for en overfart.

Løsning:

Forventet fortjeneste $E[F]$ på brussalget til *Fjord1* for en overfart:

$$E[F] = E[(p - k)B_{\text{tot}} - k_{\text{fast}}] = (p - k)E[B_{\text{tot}}] - k_{\text{fast}} \quad (1.88)$$

I oppgaven er det opplyst at $p = 20$ NOK, $k = 6$ NOK og $k_{\text{fast}} = 175$ NOK.

I tillegg fant vi i oppgave 3i) at $E[B_{\text{tot}}] = 37.95$.

Dermed:

$$E[F] = ((20 - 6) \cdot 37.95 - 175) \text{ NOK} = 356.30 \text{ NOK} \quad (1.89)$$

■

Oppgave 4 — 20%

Helse Møre og Romsdal (Helse M&R) har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R har en mistanke om at antall utrykninger øker fra år til år. Derfor bestemmer de seg for å observere antall utrykninger i perioden 2011 – 2016 og innfører følgende variabler:

$$x_i = \text{år nr. } i \quad (i = 1, \dots, 6) \quad (1.90)$$

$$y_i = \text{antall utrykninger for år nr. } i \quad (i = 1, \dots, 6) \quad (1.91)$$

	2011	2012	2013	2014	2015	2016
x_i (år nr. i)	1	2	3	4	5	6
y_i (antall utrykninger for år nr. i)	6726	7595	8451	9350	11 233	12 543

Figur 1.8: År x_i og antall utrykninger y_i .

Et Excel-plott av dataene ovenfor er vist i figur 1.10.

Dataprogrammet Excel kan også regne ut minste kvadraters regresjonslinje. Resultatet er:

$$\hat{y} = 5226.5 + 1168.5x \quad (1.92)$$

hvor altså

$$\hat{a} = 5226.5 \quad (1.93)$$

$$\hat{b} = 1168.5 \quad (1.94)$$



Foto: Colourbox



Foto: Ståle Wattø

Figur 1.9: Ambulanse.

- a) (5%) i) Forklar *kort*, på generelt grunnlag, hva en regresjonslinje mellom variablene x og y beskriver.
- ii) På generelt grunnlag, hvordan tolker du parameteren $\hat{b}$ i en *lineær* regresjonslinje?
- iii) Spesielt, hvordan tolker du parameteren $\hat{b}$ i regresjonslinjen i lign.(1.92)?

Løsning:

i) En regresjonslinje mellom variablene x og y sier at for en gitt verdi av x så kan y estimeres/predikeres via regresjonslinjen.

ii) Parameteren $\hat{b}$ er stigningstallet for en lineær regresjonslinje. Parameteren angir estimert/predikert endring $\hat{y}$ når x endres med én enhet.

iii) Parameteren $\hat{b}$ er oppgitt i oppgaven: $\hat{b} = 1168.5$
Det betyr at antall utrykninger øker med 1168.5 per år.

formel

- b) (2.5%) Anta at den lineære trenden i antall utrykninger fortsetter helt frem til 2020.

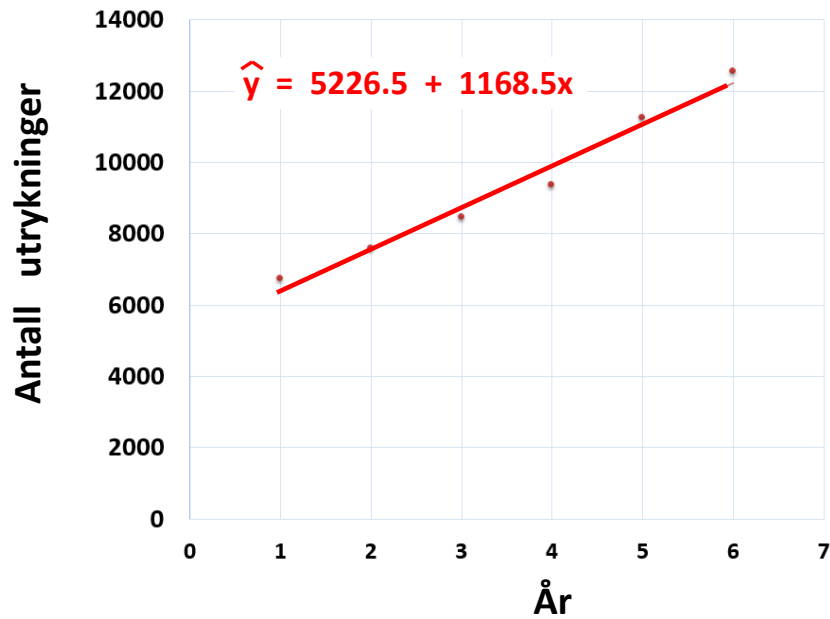
Hvor mange utrykninger predikerer regresjonslinjen at det vil være i år 2020?

Løsning:

År 2020 tilsvarer $x = 10$. Regresjonslinjen predikerer at antall utrykninger i år 2020, dvs. $x = 10$, er:

$$\hat{y}(10) = 5226.5 + 1168.5 \cdot 10 = 16\,911.5 \quad (1.95)$$

dvs. det er 16 912 utrykninger i 2020.



Figur 1.10: Plott av dataene i figur 1.8.

- c) (5%) Vis at korrelasjonskoeffisienten r_{xy} er gitt ved

$$r_{xy} = \hat{b} \frac{s_x}{s_y} \quad (1.96)$$

Løsning:

Korrelasjonskoeffisienten r_{xy} er definert ved: (se formelsamling)

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (1.97)$$

Definisjonen av koeffisienten $\hat{b}$ i regresjonslinjen er: (se formelsamling)

$$\hat{b} = \frac{s_{xy}}{s_x^2} \quad (1.98)$$

og løser denne med hensyn på s_{xy} alene:

$$s_{xy} = \hat{b} s_x^2 \quad (1.99)$$

Setter lign.(1.99) inn i lign.(1.97):

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (1.100)$$

$$= \frac{\hat{b} s_x^2}{\cancel{s_x} s_y} = \hat{b} \frac{s_x}{s_y} \quad , \quad \text{q.e.d.} \quad (1.101)$$

Ut fra tabellen i figur 1.8 kan man regne ut gjennomsnittet $\bar{x}$:

$$\bar{x} = 3.5 \quad (1.102)$$

Man kan også regne ut det empiriske standardavviket til x , dvs. S_x , og det empiriske standardavviket til y , dvs. S_y . Resultatene er:

$$s_x = 1.87, \quad s_y = 2213.9 \quad (1.103)$$

(Størrelsene i lign.(1.102) og (1.103) trenger du ikke å regne ut. Bare ta dem for gitt).

- d) (2.5%) Regn ut korrelasjonskoeffisienten r_{xy} for observasjonene ved å bruke lign.(1.96).

Løsning:

Korrelasjonskoeffisienten r_{xy} for observasjonene blir da:

$$r_{xy} = \hat{b} \frac{s_x}{s_y} \quad (1.104)$$

$$= 1168.5 \frac{1.87}{2213.9} = 0.99 \quad (1.105)$$

- e) (2.5%) Tolk svaret du fikk for r_{xy} i oppgave 4d). ¹⁷

Løsning:

Siden $r_{xy} = 0.99$ så er det et sterk positiv korrelasjon mellom x og y .
Det er en nesten perfekt lineær sammenheng mellom x og y , med positivt stigningstall.

¹⁷Hva sier den numeriske (tallverdien) av r_{xy} fra oppgave 4d) om graden av korrelasjon mellom observasjonene x og y ?

- f) (2.5%) Finn gjennomsnittlig antall utrykninger per år $\bar{y}$ uten å bruke definisjonen av gjennomsnitt.

Løsning:

Definisjonen av koeffisienten $\hat{a}$: (se læreboken)

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (1.106)$$

og denne med hensyn på $\bar{y}$ alene

$$\bar{y} = \hat{a} + \hat{b}\bar{x} \quad (1.107)$$

$$= 5226.5 + 1168.5 \cdot 3.5 = 9316.3 \quad (1.108)$$

som er gjennomsnittlig antall utrykninger per år når man tar gjennomsnittet over perioden 2011 – 2016.

■